DOCUMENT RESUME

ED 229 670 CG 016 638

AUTHOR Banks, Cristina G.
TITLE Behavioral Indices of Raters' Cognitive Processing in Performance Appraisal.
SPONS AGENCY Texas Univ., Austin. Research Inst.
PUB DATE Aug 82
GRANT R-374
NOTE 29p.; Paper presented at the Annual Convention of the American Psychological Association (90th, Washington, DC, August 23-27, 1982).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Behavior Patterns; Behavior Rating Scales; *Cognitive Measurement; *Cognitive Processes; Decision Making; Evaluation Criteria; *Measurement Techniques; Performance Factors; Research Methodology; Test Reliability; Test Validity
IDENTIFIERS Instantaneous Report of Judgments; *Performance Appraisal

ABSTRACT

One approach to increasing our understanding of the rating process is to examine behavioral components of decision-making. Although observable rater behavior during appraisal is still removed from the actual contents of internal processing, these behavioral indices may provide important clues toward identifying determinants of rating success. To measure rater behavior during appraisal a methodology called Instantaneous Report of Judgments (IRJ) was developed. Rating behaviors which are believed to reflect important dimensions of rating ability were examined: (1) amount of information utilized; (2) sensitivity to differences between ratees; (3) sensitivity to ratee strengths and weaknesses; and (4) observational style. Several studies were conducted using the IRJ. The first set consisted of basic descriptive studies of rater behavior during the rating process with the goal of identifying stable components of rating style. The second set involved construct validation of the IRJ procedure and rating data. Analyses of results from these studies showed that IRJ can provide reliable and valid data and that these behavioral indices shed some light on the underlying mechanisms of accuracy. (Author/WAS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED229670

BEHAVIORAL INDICES OF RATERS' COGNITIVE PROCESSING IN PERFORMANCE APPRAISAL

Cristina G. Banks

Department of Management

The University of Texas at Austin

Presented as part of a symposium entitled, "Methods of Investigating the Rating Process," at the annual meeting of the American Psychological Association, Washington, D.C., August, 1982.

This research was partially supported by a Unversity Research Institute Grant R=374 from The University of Texas. The author is indebted to the valuable assistance of Chet Miller and Kathy Epps who helped prepare this manuscript. Special thanks go to Robin Eddy for technical assistance in equipment development.

CG 016638

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Cristina G. Banks

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Behavioral Indices of Raters' Cognitive Processing

in Performance Appraisal

ABSTRACT

One approach to increasing our understanding of the rating process is to examine behavioral components of decision-making. Although observable rater behavior during appraisal is still removed from the actual contents of internal processing, these behavioral indices may provide important clues toward identifying determinants of rating success. A methodology called Instantaneous Report of Judgments (IRJ) was developed to measure rater behavior during appraisal. Four rating behaviors were examined which are believed to reflect important dimensions of rating ability: amount of information utilized, sensitivity to differences between ratees, sensitivity to ratee strengths and weaknesses, and observational style. Two sets of studies were conducted using IRJ. The first set consisted of basic descriptive studies of rater behavior during the rating process with the goal of identifying stable components of rating style. The second set involved construct validation of the IRJ procedure and rating data. Results from these studies are presented and discussed briefly. It is concluded that IRJ can provide reliable and valid data and that these behavioral indices shed some light on the underlying mechanisms of accuracy.

Behavioral Indices of Raters' Cognitive Processing in Performance Appraisal

Performance rating research is in a state of transition. Instead of searching for ways of improving the mechanics of appraisal (e.g., better rating forms, more time allotted to the task, more effective rater training), researchers recommend investigating the processes underlying performance rating (Feldman, 1981; Ilgen & Feldman, in press; Landy & Farr, 1980). While many have stressed the importance of this kind of research for several years (e.g., Borman, 1979, Dunnette & Borman, 1979), few studies have been completed. One reason for this delay is the absence of easy methodologies for studying process variables for psychologists in general and I/O psychologists in particular. Thus, many I/O psychologists who are interested in rating process research have borrowed paradigms outside I/O and adapted them to the appraisal context. This paper describes a methodology adapted from cognitive psycholoy for analyzing the rating process. This methodology, called Instantaneous Report of Judgments (IRJ), yields behavioral indices of raters' cognitive processing in performance appraisal.

Background and Rationale

Why examine behavioral components of the rating process? Two reasons come to mind. First, it is generally accepted that previous attempts to increase accuracy by examining the relationship between

input variables (i.e., training, experience, interpersonal accuracy correlates) and appraisal outcomes (i.e., errors) have yielded disappointing results. To help explain the failure of previous studies, we must look deeper into the relationship beween input and utcome variables. Why don't rating formats aid raters' decision-making? Why have training programs failed to improve substantially rater accuracy? A description of rater behavior (e.g., processing of information) during appraisal will help determine how various input variables affect rating process and hence, outcomes. Second, several interesting research questions may be answered by analyzing rater behavior at the micro level such as: Does sex or race bias enter into performance rating at the selection or evaluation stage of processing? Do raters search for disconfirmatory information once a judgment is formed? Do raters utilize the same information to generate performance ratings for several dimensions? In sum, knowledge of rater behavior during the rating process may suggest what to change behaviorally to increase rater accuracy.

In this paper, I will describe IRJ and present findings from studies employing the IRJ procedure. I hope to show that important new information about the rating process can be obtained through this methodology and that insights into the determinants of rating accuracy are likely by following this approach. First, let me be clear about what is meant by the term, rating process, and what constructs I intend to measure through IRJ.

The rating process is conceptualized as a five-step information-processing sequence that results in an overall performance

rating for a particular performance dimension. The steps consist of internalizing task requirements, selecting relevant information, evaluating selected information, storing and recalling stored information, and combining evaluations (See Banks, 1981 for more detail). It is important to note in this conceptualization that the task as given may not be identical to its interpretation and that information search and selection is a central component of the process. Both of these aspects of the conceptualization are important because they allow for individual differences in the selection and interpretation of ratee data, a consideration that is downplayed or ignored in other investigations of the rating process (e.g., policy-capturing). A methodology is desired that measures these individual differences explicitly because it is believed that these individual differences will play a key role in unraveling the mystery of accurate rating.

Based on this conceptualization of the rating process and on folk knowledge of the secrets of successful rating in the literature, four constructs were hypothesized to comprise rating ability: (1) degree of information utilization; (2) sensitivity to differences between ratees; (3) sensitivity to ratee strengths and weaknesses; and (4) global vs. specific observational style. These constructs are described below.

1. Degree of information utilization. This construct is defined as the amount of information a rater utilizes during a rating task. Utilization of information is considered important because the literature suggests that the more information a rater uses, the higher

the probability job-related information will influence evaluation (Schmitt, 1976).

2. Sensitivity to ratee differences. This construct is similar to one of Cronbach's components of judgmental accuracy, differential elevation (DE; Cronbach, 1955). This construct reflects a rater's ability to detect differences between ratees when differences actually exist. The higher the variance in performance ratings across ratees, the more differences a rater detects. The literature claims that a lack of differentiation, or restriction of range, leads to lower accuracy (cf. Carroll & Schneier, 1982). While this literature is based on summary or overall ratings rendered for a ratee, sensitivity to ratee differences could be extended to the level of individual judgments which compose summary ratings.

3. Sensitivity to ratee strengths and weaknesses. This construct attempts to capture a rater's ability to evaluate performance in an even-handed or balanced manner. Within a performance dimension, a lack of sensitivity has been characterized as a failure to seek or recognize disconfirmatory ratee information after an impression is established (Snyder & Swann, 1978). A confirmatory strategy ,one in which a rater seeks information consistent with his or her impression, may result in low variability in information utilized, and hence, failure to utilize all relevant information if both positive and negative information are present.

4. Global vs. specific observational style. This construct attempts to capture the kind of information a rater processes during appraisal. "Global" processors may be characterized as those who

develop global impressions of the ratee by processing information at a more abstract level than "specific" processors. Global processors do not develop impressions on the basis of specific behavioral events; rather, they form impressions by generalizing across ratee behaviors, forming abstractions from the behavioral data. For example, a global processor may attend to the ratee's attitude across all incidents that involve a conflict with a subordinate. In this case, the rater may be evaluating a performance dimension different from the one explicitly stated on the rating form (e.g., "attitude" vs. "ability to resolve conflict"). It is believed that specific processors, on the other hand, generate summary ratings by combining separate and specific bits of information and avoid generalization across incidents. This latter style may reduce the probability a few salient events will swamp subsequent judgments (Schmitt, 1976).

Each construct was operationalized by an observable rating behavior emitted during the rating process. Constructs and associated rating behaviors are listed in Figure 1. Notice that the amount and kind of information utilized requires that one know the number and content of judgments made by a rater. Instantaneous Report of Judgments (IRJ) was developed so that raters could describe their judgments when they occured during a rating task. A description of IRJ and how these constructs were measured follows.

Insert Figure 1 about here

Instantaneous Report of Judgments

Instantaneous Report of Judgments or IRJ was based for the most part on information processing theory as presented in cognitive psychology (Ericsson & Simon, 1980; Newell & Simon, 1972). Briefly, a rater reports his or her judgments formed during a rating task by using a panel of buttons to record judgments of ratee performance and by reporting verbally behavioral cues that "trigger" judgments. (See Banks, 1980 & 1981 for more detail.) Basically, IRJ provides raters a mechanism for reporting the contents of their decision-making whenever they feel the "urge" to report.

The four behavioral indices of raters' cognitive processing (number of judgments, variation in judgments, variation in mean judgments, and latency) are obtained in the following way. Each button press on the panel signals a judgment was made; therefore, the number of button presses indicates the number of judgments made (NJ). Since button values duplicate the point values on the rating scale, the particular button pressed indicates the judged level of ratee performance. Variation in judgments is obtained by the standard deviation of the values of buttons pressed (SDJ). When these values are averaged yielding a mean judgment level per ratee, variation in mean judgments is obtained by calculating the standard deviation of mean judgments across ratees (SDJ). A timing device which ties button presses to on-going ratee behavior allows measurement of latency (LAT). It also ties judgments to ratee cues, permitting identification of information utilized by a rater in forming a

judgment. Thus, IRJ allows measurement of four rating behaviors, plus identification of cues selected and processed during the rating task. These operations, in turn, allow measurement of the constructs believed to be related to rating ability.

Raters in IRJ studies individually view videotaped performances of managerial behavior (5-7 minutes long). Videotaped were previously developed by Borman and his associates (Borman, Hough, & Dunnette, 1976). Raters view and rate a single performance dimension for each manager. In other words, one manager is evaluated on one dimension per viewing, and this constitutes a single rating task. In each rating task, raters press a button whenever they "feel" they are making a judgment, and they press the button (1-7) that best represents their judgment of ratee performance. After pressing a button, they report verbally the basis for their judgment. For every task, raters are encouraged to press buttons as many times as they make judgments and at the conclusion of each task, they render a summary rating. In all, six ratees are rated along each of six performance dimensions.

Research Findings

Several studies have been conducted using IRJ, and these are outlined in Figure 2. These studies can be divided into two

---

Insert Figure 2 about here

---

groups: descriptive studies of the rating process and construct validation of IRJ. The descriptive studies were designed to collect

basic information about rater behavior during a rating task. They sought to determine how much information raters utilize and what information is utilized, and to determine the presence (absence) of a general rating style. Details of these studies can be found in Banks (1981; 1982). Findings of these studies will be summarized briefly below.

In terms of judgments made, a rater makes about seven judgments per ratee, though large individual differences exist in the number. A rater tends to make judgments early in the evaluation period (within 2 minutes), and the range of judgments made for each ratee is relatively small (within 1 to 2 points on a 7-point scale). A rater also does not differentiate greatly across ratees; the range of mean judgments is about 1.5 points. When rating behavior is observed across tasks, marked similarities in rating behavior were found. This suggests that a rater tends to utilize a consistent rating style across tasks. For example, raters appeared to be consistent regarding number of judgments made (NJ) and judgment latency (LAT), but variation in judgments (SDJ) was less consistent (median internal consistency reliabilities = .95, .77, and .61, respectively). An interesting finding emerged when SDJ was examined across tasks. This is, raters seemed to narrow their range of judgments with practice. It is not clear whether this narrowing of judgments was the result of becoming more skilled over time or whether experience with the task changed their reporting.

When cue selection and evaluation was examined, it was found that untrained raters do not tend to select the same information when they

evaluate the same ratee along the same performance dimension. Moreover, even when raters selected the same information, they evaluated it differently. These latter findings suggest that untrained raters (college students) differ substantially in the factors affecting information or cue selection (e.g., interpretation of task requirements, motivation, attention) and cue evaluation (e.g., rating criteria, cognitive schema, preconceived notions of ratee performance). Simply providing well-developed rating formats like BARS and removing conflicting motives (e.g., eliminating responsibility for the ratings) is not sufficient to guide the rating process to the same end.

The second set of studies sought to determine the meaningfulness of these rating behaviors. First, a rate-rerate reliability study was conducted to determine if these rating behaviors were repeatable when identical tasks were administered 1 t 5 months later. For a subset of 16 raters, mean judgments calculated for Time 1 and Time 2 tasks were highly correlated as were overall performance ratings (median r's = .82, .83, respectively). These findings suggest that a rater arrived at the same outcome at both administrations. For NJ, LAT, and especially SDJ, reliability was lower (median r's = .54, .49, and -.05, respectively). A rater tended to press a different number of buttons (usually fewer) and pressed a smaller range of buttons upon the second viewing. As with internal consistency analyses, this analysis suggests some revision in rating behavior with practice, thus lowering reliability estimates. But, one could argue that the rating tasks were no longer identical since a rater possessed more

information about the ratee in the second viewing than the first. This would result in artificially low estimates of reliability. Overall reliability analyses suggest that although judgments differ in quantity and range over time, they combine to form the same conclusion, a finding that argues against the possibility that raters responded randomly.

Generalizability of IRJ findings was assessed in part by comparing managers' with students' rating behaviors in identical rating tasks. Both managers and students rated each of the six ratees along each performance dimension in a total of six rating sessions. Managers and students were compared in terms of rating behavior (NJ, SDJ, SDJ, and LAT), and rating outcomes (accuracy, halo, leniency, and restriction of range). Various person perception variables shown to be related to rating success (Borman, 1979) were also compared. Means and standard deviations of rating behaviors, rating outcomes, and person perception variables are shown in Table 1.

Insert Table 1 about here

No significant differences were found between the two groups except for age and cognitive complexity (students were younger, but smarter). Although managers and students do not differ significantly on these variables, some important pattern differences in the relationships between variables were evident. Pattern differences will be elaborated on in a later section. For the moment, let us examine each variable singly. In general, the behavior of managers and students in

the rating tasks was quite similar, suggesting that we would not expect managers in general to respond very differently when given the same tasks as students.

Another study was conducted to determine whether reporting raters' contents of their decision-making altered rating outcomes. If so, this would limit the generalizability of IRJ findings to typical rating tasks. Mean performance ratings were calculated across raters from The Banks (1979) sample for each ratee on each dimension. These mean ratings were correlated with mean performance ratings collected by Borman (1979). Borman's ratings were obtained by having raters simply view the same videotapes and record summary performance ratings. Despite differences in procedure, samples, and rating instructions, ratings from the two studies correlated .90 ($p<.01$) and the sum of the differences between the two groups of means was near zero ($d = .3$). Similar correlations with the Borman ratings were found with mean ratings from a later IRJ study ($r = .91$, $p<.01$). Recently, I collected ratings from an independent sample of student raters ($N = 37$) using the Borman procedure, and again the correlations between these ratings and ratings from the IRJ studies were high (r's = .94 & .96). In general, it can be concluded that findings from IRJ studies can probably be generalized to rating tasks typically encountered in appraisal research. More important, the IRJ procedure does not seem to interfere greatly with the rating process.

Next, rating behaviors collected using IRJ were correlated with rating outcomes (accuracy, halo, leniency, and restriction of range) to determine which rating behaviors were associated with accuracy and

rating error. Data from managers and students were examined separately. Originally, the manager sample was separated into two subgroups, expert and nonexpert raters. Experts were differentiated from nonexperts on the basis of textbook-type knowledge of appraisal and on the basis of rating experience as judged by the author after in-depth interviews. These subgroups were combined when no significant differences in rating behaviors or rating outcomes were found. (So much for armchair criterion analyses.) Relationships between rating behaviors and rating outcomes for each group are shown in Table 2. It can be seen that restriction of range

---

Insert Table 2 about here

---

error is consistently related to AVGSD, the variation in mean judgments (SDJ) averaged across ratees. AVGSD is the micro-level analog of restriction of range since both are computed in terms of differentiation between ratees. A high correlation between the two measures means that differentiation (or lack of it) at the judgment level is consistent with differentiation at the summary rating level. For the manager sample, leniency was related to AVGNJ, the number of judgments (NJ) averaged across tasks, and AVGSD, the average variation in mean judgments (SDJ). This suggests that the more judgments a rater makes and the greater the differentiation between ratees, the lower the leniency. However, these correlations were not found in the student sample.

The most interesting aspect of Table 2 is the relationship between rating behaviors and accuracy. Accuracy was measured by the

correlation between each rater's set of 36 summary ratings and Borman's mean expert ratings (Borman, 1979). For the student sample only, accuracy was related to AVGNJ, AVGSDJ, and AVGLAT, aggregate scores of NJ, SDJ, and LAT averaged across tasks. These correlations showed that accurate raters tend to make fewer judgments, exhibit less variation in judgments, and take more time generating the first judgment than less accurate raters.

At first glance, correlations between rating behavior and accurancy appear to contradict expectations set forth earlier in this paper. Recall that it was hypothesized that rating accuracy would be associated with high NJ, high SDJ, and low LAT, according to the performance appraisal literature. In the student sample, the opposite seemed to be true; raters who made few judgments and exhibited longer latencies of responding tended to be more accurate. This potential inconsistency can be explained by exploring the process by which judgments are produced.

Early responders could be responding appropriately or inappropriately depending on the cues responded to. If cues relevant to the evaluation of a particular performance dimension are present early in the ratee performance, a quick response would be expected and appropriate. However, if the raters responds early to irrelevant cues (a sign of failure to discriminate cues), then early responding would be inappropriate. It may also be the case that even if relevant cues are responded to early in the process, the rater may fail to report judgments until a sufficient amount of confirmatory (or disconfirmatory) evidence has accumulated to build confidence in the

judgment. Since raters in the student sample were relatively inexperienced in giving performance appraisals, it would seem reasonable to hypothesize that their inexperience led to cautious (and therefore delayed) reporting for accurate raters and more spontaneous reporting for less accurate raters.

The lack of correlations for the manager sample also seems disturbing, but this too can be explained. The scatterplot of the relationship between accuracy and each rating behavior revealed moderate curvilinear relationships. For NJ, accurate raters tended to make either a high number of judgments or a low number whereas less accurate raters made about an average number (eta = .39). Since the eta coefficient (.39) is higher than the Pearson coefficient (r=.01) between NJ and accuracy, we can conclude that a nonlinear relationship does indeed exist. These data suggest that acccurate raters in the manager sample exhibit one of two styles of responding: early responders who have the experience and confidence to identify and report relevant cues and late responders who wait for evidence to accumulate before reporting. Thus, accurate managers may be characterized as exhibiting one of two styles of rating whereas accurate students exhibit only one. Students may not be sophisticated enough in appraisal to have developed a fine-tuned cognitive schema for interpreting performance-related cues confidently or for recognizing subtle behavioral cues. The "gist" of the cues may be obvious to the students in the aggregate, but taken singly, cues may not be interpreted as well as they would by managers. At this point, this explanation for the research findings should be regarded as

speculative.

Finally, person perception variables were correlated with rating behavior and outcome measures. These correlations are found in Table 3. As in Borman's (1979) study, intellectual factors were correlated with accuracy, but for managers only. (Correlations in the student sample may have been artificially low due to restriction of range in intellectual ability.) When rating behaviors were correlated with person perception variables, a different pattern of correlations were found for managers and students. For managers, many person perception variables

---

Insert Table 3 about here

---

were related to AVGNJ and AVGLAT whereas for students, few correlations were found. This suggests that managers' rating behavior (NJ and LAT) was affected by person perception variables more than they were for students. A more interesting finding is that although AVGNJ and AVGLAT were related to appraisal knowledge, near-zero correlations were found between appraisal knowledge and accuracy for both groups. Apparently, what raters actually do to achieve accuracy is not necessarily what textbooks suggest. This observation is worth dwelling on a minute. Originally it was hypothesized that the more information a rater utilizes and the quicker he or she responds, the higher the accuracy. While rating behaviors were consistent with knowledge of "good" appraisal techniques, they were not correlated as hypothesized to accuracy. These findings imply that our folk knowledge of "good appraisal" may be inaccurate and that we need to

rethink what rating behaviors would be expected to result in accurate rating.

Two research studies underway concern raters' use of cues. The first examines whether a rater uses the single behavioral cues for evaluating more than one performance dimension. An earlier study which used a between-subjects design showed that some behavioral cues were salient across several dimensions (Banks, 1982). If a single rater utilizes the same cue across dimensions, then halo "error" may be reintepreted as a by-product of normal decision-making rather than the result of overgeneralized global impressions. That is, halo would be caused by the overlap in information used to generate dimension ratings. If halo error is in fact a problem of multiple-cue use, training programs to reduce halo error may be more successful if raters are trained to increase their reliance on more discriminating cues.

The second study in progress involves raters' identification of relevant cues. It is essentially a study of raters' ability to separate relevant from irrelevant information. It is expected that those who utilize a high proportion of relevant information (to total information utilized) will be more accurate.

In summary, a good deal of descriptive work on the rating process has been completed. We found that raters exhibit a rating style that is consistent across tasks in many respects, but some revision in this style occurs with practice. We also found that accurate managers exhibit two different rating styles whereas accurate students exhibit only one. Reasons for these differences between styles were explored.

Nonetheless, in each case, specific rating behaviors were related to rating accuracy (though opposite to expectation). Finally, it was found that while folk knowledge of "good" appraisal techniques was related to rating behaviors in the hypothesized direction, appraisal knowledge failed to correlate significantly with rating accuracy. The correlations between rating behaviors and rating accuracy and the lack of correlation between appraisal knowledge and accuracy suggest that we need to revise our thinking about what kind of rating behavior is related to accuracy. And last, we concluded that the IRJ procedure does not seem to interfere with raters' cognitive processing and that IRJ yields for the most part, reliable and valid data.

Although a good deal of work is completed, more remains. This paper intended to show that rating process studies can be done, though slowly. This work suggests to me that the rating process is quite complex and fraught with potential errors. Knowledge gained from such work has opened up new avenues of thinking about appraisal and how to reduce potential errors. Let me elaborate on that point.

Typical appraisal systems apparently require raters to be good test developers. The only parts of the "test" a rater is given to measure a subordinate's work performance are the constructs to be measured and definitions of those constructs (with some hints as to what items may be relevant—they are called behavioral anchors or examples). Raters are left with the problem of figuring out what items (sic. behaviors) should be observed to evaluate performance and what their discriminating power is, how items should be scored and combined, and finally how raw scores should be interpreted. No wonder

inexperienced, untrained raters are unmotivated to do it when they do it poorly and are held accountable! Taking "test development" out of the appraisal process may improve rater accuracy. Training in assessment similar to assessor training in assessment centers is another possible change. I will leave the reader to think of others.

In conclusion, I believe we need to push ahead with rating process research to learn what variables affect the rating process and more important, which lead to accuracy so that we can begin to design specific and potent interventions.

Footnotes

[1] The literature's recommendations are considered folk knowledge since they are unproven but believed.

[2] Since NJ is correlated with LAT .79 for both samples, they will be used interchangeably in the discussion.

Table 1

Means and Standard Deviations of Rating Behaviors, Rating Outcomes, and Individual Difference Variables for Managers and Students

	MANAGERS		STUDENTS		F	Significance
VARIABLES	$\bar{X}$	SD	$\bar{X}$	SD		
AVGNJ	3.78	1.86	3.58	1.97	NS	NS
AVGSDJ	.69	.24	.75	.28	NS	NS
AVGSD	.19	.04	.18	.03	NS	NS
AVGLAT	183.74	116.89	206.16	98.6	NS	NS
Halo	1.08	.35	1.19	.32	NS	NS
Leniency	3.60	.44	3.72	.42	NS	NS
Restriction of Range	2.09	.37	2.02	.27	NS	NS
Accuracy	1.03	.27	1.16	.28	NS	NS
Embed Figures	13.08	4.18	14.95	2.70	2.39	.046
Bieri Cognitive Complexity Form	94.27	25.13	89.15	11.52	4.76	.001
Age	32.66	14.31	23.15	8.06	3.15	.01
Detail Orientation	9.50	3.11	9.70	2.79	NS	NS
Task Orientation	8.50	3.21	8.40	3.54	NS	NS
Intellectual Ability and Interest	5.55	2.63	7.00	2.44	NS	NS
Personal Adjustment	36.27	12.34	36.95	9.59	NS	NS
Realistic Theme	2.33	1.51	2.50	1.39	NS	NS
Investigative Theme	2.44	1.53	2.80	1.88	NS	NS
Artistic Theme	2.83	1.76	3.20	1.70	NS	NS
Social Theme	4.16	1.69	3.90	1.77	NS	NS
Enterprising Theme	4.63	1.86	4.70	1.75	NS	NS
Conventional Theme	2.75	1.81	2.85	1.72	NS	NS
Outgoing vs. Shy	6.22	2.52	6.35	2.32	NS	NS
Adjusted vs. Malady	6.86	2.40	7.00	2.61	NS	NS
Decisive vs. Indecisive	6.41	2.41	6.45	2.25	NS	NS
Friendly vs. Unfriendly	6.94	2.70	7.65	1.95	NS	NS
Interested in Others vs. Self-Absorbed	6.88	2.40	7.20	2.09	NS	NS
Cheerful vs. Humored	6.58	2.60	6.95	2.06	NS	NS
Dominant vs. Submissive	5.88	2.57	6.55	2.28	NS	NS
Considerate vs. Inconsiderate	6.75	2.61	7.50	1.96	NS	NS
CPI Tolerance Score	13.05	5.05	12.95	3.54	NS	NS
CPI Well-Being Score	18.66	6.58	18.35	5.63	NS	NS
CPI Stress Reduction Score	4.55	4.37	5.65	3.82	NS	NS
Highest Education Level	3.75	2.07	3.55	1.60	NS	NS
High School GPA (5 pt. scale)	3.11	1.48	4.20	1.19	NS	NS
Importance of Appraisal Procedures	71.25	23.58	62.10	27.76	NS	NS
Appraisal Knowledge Test	12.41	4.67	13.25	4.15	NS	NS

Table 2

Correlations Between Rating Behaviors and Rating Outcomes for Managers and Students

Rating Behaviors	Rating Outcomes			
	Halo	Leniency	Rest. Range	Accuracy
MANAGERS				
AVGNJ	.16	-.35**	.18	.01
AVGSDJ	.15	-.05	0	-.01
AVGLAT	.09	.24	-.15	-.12
AVGSD	.12	.37**	.84***	-.03
STUDENTS				
AVGNJ	-.11	.10	.22	-.53**
AVGSDJ	-.05	.36	-.22	-.43**
AVGLAT	.23	-.13	.04	.46**
AVGSD	.13	.01	.66***	.13

*p<.05
**p<.01
***p<.001

TABLE 3

Significant[b] Correlations Between Rating Behaviors and Individual Difference Variables for Managers and Students

Individual Difference VARIABLES[a]	Rating Behaviors							
	Managers				Students			
	AVGNJ	AVGSDJ	AVGLAT	AVGSD	AVGNJ	AVGSDJ	AVGLAT	AVGSD
EMBF								
CC								
DET	.40		-.48					
TA	.35		-.42					.40
IA			-.35					
PA	.33		-.44					
RT								
IT			-.32					-.42
AT								
ST	.32		-.35		.41			.40
ET			-.29			-.51		.48
CT								
OUT						-.38		.44
ADJ	.32		-.32					
DEC								
FRIEND			-.37					.37
OTHERS	.28		-.34					.38
CHEER								.47
DOM								.38
CONSID			-.29					
TOL	.29		-.44					
WB			-.32					
CR	.35				.41		-.48	
HED	.45		-.46					
GPA			-.30					
IMP	.40		-.44					.38
APP TEST	.46		-.49					

[a]See Table 1 for complete names of variables listed in this table.

[b]Correlations are reported if $p<.05$ or greater.

References

Banks, C. G. A laboratory study of the decision-making processes in performance evaluation. Unpublished doctoral dissertation, University of Minnesota, 1979.

Banks, C. G., IRJ: A new technique for measuring the performance rating process. Proceedings of the Military Testing Association, Toronto, Canada, 1980.

Banks, C. G. Instantaneous report of judgments: A technique for analyzing the performance rating process. Department of Management Working Paper No. 81/82-4-6, The University of Texas, Austin, TX, 1981.

Banks, C. G. Cue selection and evaluation elicited during the rating process. Department of Management Working Paper No. 81/82-4-11, The University of Texas, Austin, TX, 1982.

Borman, W. C., Individual differences correlates of accuracy in evaluating others' performance effectiveness. Applied Psychological Measurement, 1979, 3, 103-115.

Borman, W. C., Hough, L. M., & Dunnette, M. D. Performance ratings: An investigation of reliability, accuracy, and relationships between individual differences and rater error. Personnel Decisions, Inc., Minneapolis, Sept., 1976.

Carroll, S. J. & Schneier, C. E. Performance appraisal and review systems: The identification, measurement and development of performance in organizations. Scott, Foresman, & Co., Glenview, Illinois, 1982.

Cronbach, L. J. Processes affecting scores on 'understanding others' and 'assumed similarity.' Psychological Bulletin, 52, 1955, 177-193.

Ericcson, K. A. & Simon, H. A. Verbal reports as data. Psychological Review, 1980, .

Feldman, J. M. Beyond attribution theory: Cognitive processes in performance appraisal. Journal of Applied Psychology, 1981, 66, 127-148.

Ilgen, D. R. & Feldman, J. M. Performance appraisal: A process focus. In L. L. Cummings and B. M. Staw (Eds) Research in Organizational Behavior, Vol. 5 (in press).

Landy, F. J. & Farr, J. L. Performance rating. Psychological Bulletin, 1980, 87, 72-107.

Newell, A. & Simon, H. A. Human problem solving. New Jersey: Prentice-Hall, Inc., 1972.

Schmitt, N. A. Social and situational determinants of interview decisions: Implications for the employment interview. Personnel Psychology, 1976, 29, 79-101.

Snyder, M. A. & Swann, W. B., Jr. Hypothesis testing process in social interaction. Journal of Personality and Social Psychology, 1978, 36, 1202-1212.

Figure 1

Behavioral Constructs, Operations, and Variable Names

CONSTRUCT	OPERATION	VARIABLE
Degree of Information Utilization	Number of judgments made for for each ratee	NJ
Sensitivity to Differences Between Ratees	Variation in mean judgments for each ratee	$SD\bar{J}$
Sensitivity to ratee Srengths and Weaknesses	Variation in judgments for each ratee	SDJ
Observational Style	Latency before first judgment	LAT

Figure 2

Research Completed/In Progress

I. Descriptive Studies of Rating Behavior

Study 1. Number and Kinds of Judgments

Study 2. Cue Selection and Evaluation

Study 3. Stability of Rating Behavior Across Ratees

II. Construct Validation of IRJ

A. Robustness of the Technique

Study 5. Generalizability of IRJ Results

Study 4. Rate-rerate Reliability of Rating Behavior

Study 6. Impact of Reporting

B. Validation of Behavioral Data: Correlations with Various Rating Outcomes

Study 7. Rating Behavior and Rating Outcomes

Study 8. Rating Behavior and Correlates of Accuracy

(In Progress) Study 9. Multiple Cue Use and Halo Error

(In Progress) Study 10. Identification of Relevant Cues and Accuracy